What processes must we understand to forecast regional scale population dynamics?

Jesse R. Lasky[1], Mevin B. Hooten[2], and Peter B. Adler[3]

[1]Pennsylvania State University, Department of Biology
[2]U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit; Colorado State University, Departments of Fish, Wildlife, & Conservation Biology and Statistics
[3]Department of Wildland Resources and the Ecology Center, Utah State University

December 9, 2020

Abstract

An urgent challenge facing biologists is predicting the regional scale population dynamics of species facing environmental change. Biologists suggest that we must move beyond predictions based on phenomenological models and instead base predictions on underlying processes. For example, population biologists, evolutionary biologists, community ecologists, and ecophysiologists all argue that the respective processes they study are essential. Must our models include processes from all of these fields? We argue that answering this critical question is ultimately an empirical exercise requiring a substantial amount of data that have not been integrated for any system to date. To motivate and facilitate the necessary data collection and integration, we first review the potential importance of each mechanism for skillful prediction. We then develop a conceptual framework based on reaction norms, and propose a hierarchical Bayesian statistical framework to integrate processes affecting reaction norms at different scales. The ambitious research program we advocate is rapidly becoming feasible due to novel collaborations, data sets, and analytical tools.

1 Overview

From systems biology [1] to ecosystem ecology [2], researchers claim that 1) prediction is an acid test of our understanding of biology and 2) the best predictions will come from models rooted in the processes that generate system-wide patterns. This is a growing emphasis among organismal biologists, who often argue that prediction of population and species-level dynamics should be grounded in genetics, physiology, and community ecology [3–7].

A key challenge facing contemporary organismal biologists is predicting or forecasting regional-scale population dynamics under environmental change. We use the terms "prediction" and "forecast" interchangeably, defined as expected states of nature in the future. Reliable predictions could guide biodiversity conservation under future environments [8–10]. For example, a model capable of predicting the dynamics of a threatened or invasive species across a large region could identify future critical habitat or help managers allocate resources for eradication. These predictions have been traditionally based on site-specific population models or large-scale, phenomenological distribution models [11]. However, these approaches are frequently criticized [9, 12, 13], whereas a number of prominent recent reviews have made the case for models based on lower-level mechanisms [3, 5–7]. Phenomenological models are based on empirical correlations between environment and occurrence or abundance without explicit functions representing lower-level biological processes. In contrast, mechanistic approaches have explicit representations of how the environment interacts with genotypes, phenotypes, or demography.

But which processes are essential? Many population biologists believe that useful forecasts are impossible without accounting for demographic mechanisms [14]. Evolutionary biologists argue that models must consider genetic variation and adaptation [3, 15]. Community ecologists focus on species interactions [4, 5]. Organismal ecologists argue that physiology [6, 16–18] or behavior [19, 20] are essential. Must our models include all of these processes?

We suggest that this is an empirical question that can only be answered by making predictions with competing models and validating them with independent observations. However, testing the predictive value of so many processes requires a program of data collection and integration that, to our knowledge, has never been completed for any system. Our purpose is to motivate and facilitate such an effort. We begin by reviewing the arguments for building multiple mechanisms into forecasts of regional-scale dynamics. We then develop a conceptual framework for integrating diverse mechanisms, and propose a statistical framework to test their impact on predictive skill while accommodating uncertainty.

2 Toward a Process-Oriented Approach

2.1 Shortcomings of a phenomenological approach

Population predictions based on a purely phenomenological representation of biological systems may lack fidelity under novel conditions or when the underlying drivers are high dimensional. In the first case, the functional form of responses to the environment may change as organisms encounter conditions not previously experienced, or as individuals develop trait values and trait combinations not previously observed [21, 22]. These novel conditions may arise in space (e.g., for an invasive species) or in future times (e.g., under climate change). Note that forecasts have specific temporal scales: many ecological forecasts apply to the next few years, or to the next few decades. A consideration of novel conditions may be particularly problematic for long-term forecasts as conditions with no contemporary analog become more common.

In a variety of fields, such as epidemiology of infectious disease, fitting curves to past observations and extrapolating to predict the future has yielded poor results as conditions change [23, 24]. In principle, mechanistic models avoid these problems by capturing the processes that lead to novel responses to environment. It is important to note that we focus on approaches that both model underlying processes *and* include lower-level data on these processes. Mechanistic approaches that lack such data may suffer from problems with identifiability, for example, when high-level abundance data can be reproduced by models with very different mechanisms [25, 26]. Therefore a crucial feature of the approach we present below is that mechanistic models are combined with *data on underlying mechanisms* (Figures 2 and 3).

High dimensionality limits the feasibility of data-driven phenomenological approaches to prediction. When true functional forms are unknown and there are many potentially important and interacting independent variables, it is difficult to collect sufficient high-quality data to describe the true, complex relationships driving regional population dynamics. When multiple population drivers have opposing effects or opposing patterns of variation in nature, observational studies can easily miss these signals. This dimensionality is especially problematic when observations are made at a scale where aggregation obscures the effects of important processes [27–29]. A mechanistic understanding can identify the most important aspects of organisms, environments, and scales of observation. In the following sections, we describe examples of the potential benefits of process-oriented approaches, focusing on four categories of biological complexity that can confound regional population forecasts (Figure 1): Genetic variation, complex environment-performance relationships, biotic interactions, and demography.

2.2 Genetic variation

Genetic variation within species is often overlooked in population, community, and macro-ecology. Models and predictions of species range dynamics or regional population forecasts typically assume a single set of parameters for a species [30].

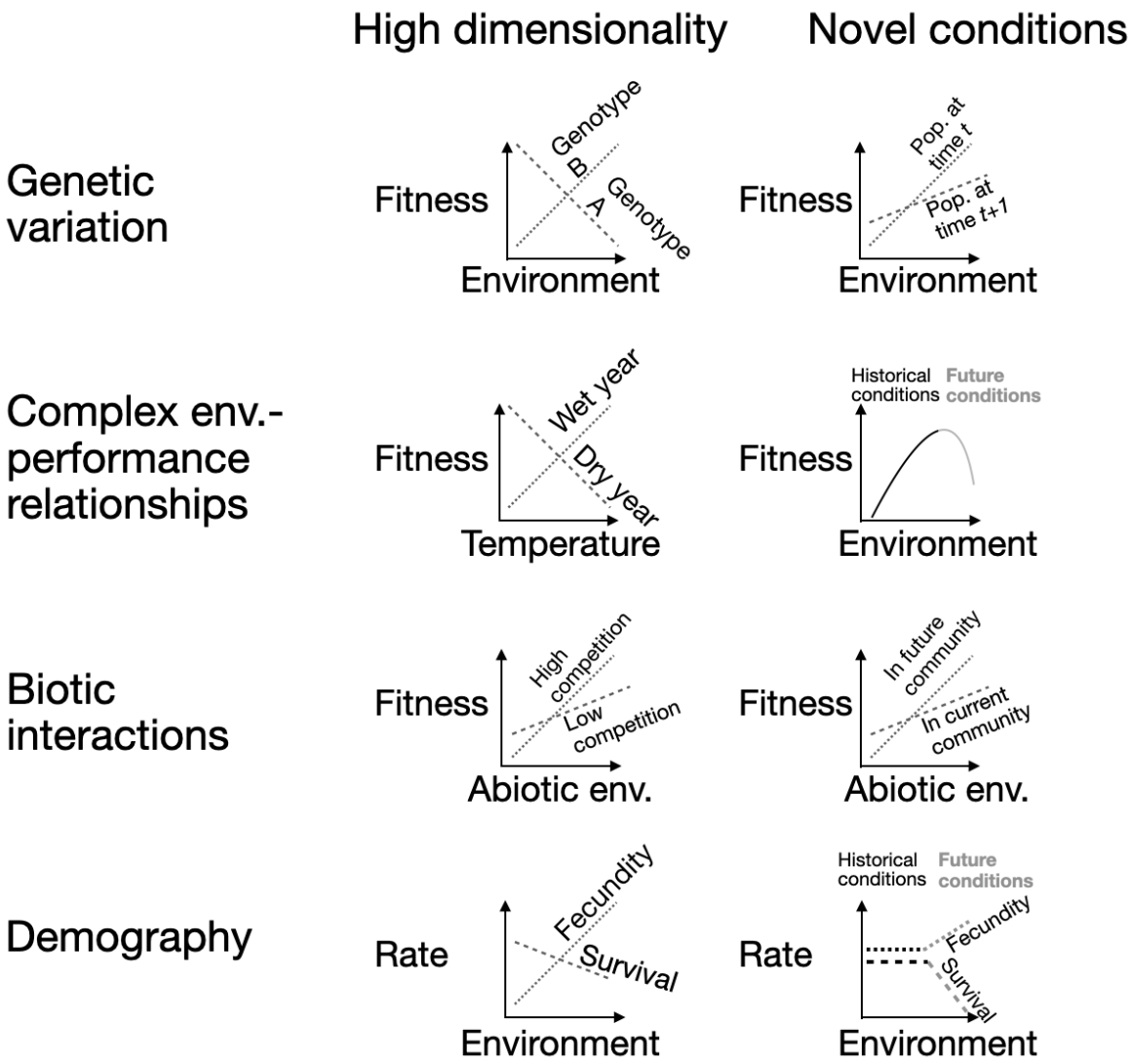

Figure 1: Examples of how underlying biological complexity may confound regional population forecasting, due to high dimensionality of the systems or the systems entering novel conditions. For each example, ignoring the complexity shown here could lead to major inaccuracies in forecasts. In the "high dimensionality" examples, relationships between environment and fitness are high dimensional due to the four different categories of processes. In the "novel conditions" examples, the relationship between environment and fitness observed in contemporary populations change under novel conditions related to the same four categories of processes.

In reality, species often exhibit intraspecific genetic variation in response to their environment. Evolutionary biology offers numerous examples of genetic variation in trait and demographic responses to environmental variation, such as for abiotic stress response [31–33], host-pathogen interactions [34], and predator-prey interactions [35, 36]. Spatially-varying environments can favor different traits in different places and, as a result, much of the genetic variation in environmental responses is geographically structured. As a result, individual fitness and population growth rates often respond to the environment in different ways in different locations. Large-scale forecasts typically ignore these differences.

If populations differ genetically in environmental responses, then accurate regional population forecasts may require accounting for genetic variation, as well as genetic change over time (i.e., evolution, Figure 1). For example, forecasts of regional population dynamics might assume individuals have as broad an environmental tolerance as the species in aggregate, while in reality locally adapted populations have narrower tolerances [30]. In this case, even though populations suited to future environments might exist within a species' range, specific alleles may not occur in the location where they would be adaptive. As a result, forecasts under environmental change might be overly optimistic and underestimate the threat of extinction [37].

Researchers have begun to incorporate genetic variation into regional population prediction. Some researchers have fit models that identify genotypes associated with specific current environments, assuming local adaptation, and then used predicted future environments to assess mismatch between local genotypes and their future environments [38–41] (reviewed by [15]). Additionally, researchers have fit environmental response models using geographically restricted subsets of populations, presumably accounting for geographic genetic variation, to project distributions under future conditions [42–44].

Without rigorous validation, predictions have unknown value. Important hurdles for predictions are success at out-of-sample prediction, and prediction into parameter space where data are sparse. After all, these models are being developed for prediction in a future of novel conditions. Recently, genetic models of environmental response and adaptation have been used in out-of-sample predictions of individual or population performance in response to environmental stressors in both wild and agricultural species [38, 41, 45–48]. These have generally had modest success, for example in predicting relative genetic variation in change in performance [46, 47] or population change [38] with accuracy (predicted versus observed correlation) ≈ 0.3. [48] were able to predict relative genetic variation in performance in novel (non-test) environments, using a combined genomic-physiological model, to $r \approx 0.6$.

Incorporating genetic variation in regional population forecasts presents several challenges. First, statistical interactions among genetic loci can be important, but may be unknown *a priori* and too high-dimensional to discover with data driven approaches. In crop systems, where there have been major advances in using genetics to predict environmental responses, models incorporating genetic variation have largely treated genetics as phenomenological [49]. Opportunities may exist to improve forecasts with largely phenomenological treatment of genetic effects, but focused on capturing some important aspects of the underlying genetic mechanism. For example, predictions may be based on subsets of genetic loci (e.g., gene regulatory loci) most likely to affect ecologically important traits and environmental responses [50–52].

Second, novel population genetic compositions will evolve through new mutations, gene flow, selection, and drift. Some of these processes are particularly challenging to model [30]. For example, gene flow is often poorly observed, and can result in major shifts in genetic composition of populations, potentially alleviating maladaptation to future environments [53]. Predictive modeling of drift may require demographic data (see below). Given the rate of global environmental change and the high standing variation often observed in traits related to abiotic stress response [54], new mutations are likely to be of lesser importance for regional dynamics (although this may not be true at expanding range margins, due to drift) [55]. Selection can be modeled using experimental data relating fitness to genotype [33], although indirect approaches may be required for species not amenable to experimentation [56, 57].

2.3 Complex environment-performance relationships

Biologists often study how individuals, populations, or species respond to a single environmental gradient, or perhaps a small set of environmental factors, along with their pairwise interactions or linear combinations [58, 59]. But interactions among environmental drivers are more complex: precipitation, temperature, and soil characteristics determine soil moisture; snow cover and air temperature determine winter soil temperature; solar radiation, albedo, wind, and air temperature determine the energy balance of ectothermic plants and animals. Additional complexity arises as organismal trait changes in response to the environment moderate the effects of potential stressors. For example, when the timing of developmental transitions (phenology) is sensitive to environmental thresholds, organisms can show sharp trait changes [60]. Animals can use behavior to avoid thresholds beyond which abiotic conditions can become dangerous [6]. As a result, individual fitness responses to environment are often nonlinear and non-monotonic, as in mortality caused by freezing temperatures, rapid nonlinear declines in performance at high temperatures, or hydraulic failure under moisture stress in plants [42, 61].

It is difficult to detect thresholds and complex interactions among measured environmental variables, especially given the noisy relationships between those variables and abundance or fitness components. The potential model space to explore is large. While machine learning approaches have been used to predict species or genotype geographic distributions based on complex environmental interactions [62, 63], they are not generative models and cannot be easily customized for specific studies, data types, and questions. Machine learning approaches can sometimes excel at prediction when they are used in the exact setting they were designed for, but they lack the ability to formally incorporate mechanisms via thoughtfully developed simulation models (e.g., mathematical models, agent-based models, virtual ecologist models; [64]). Furthermore, nonlinearities that only emerge under novel conditions may thwart extrapolations of phenomenological relationships. Forecasting models that do not account for thresholds and interactions might not predict large changes in populations moving across thresholds under future environments (Figure 1).

Direct observation of individual responses to changing environments, especially under controlled conditions, may provide more precise information for building this complexity into predictive models [19, 65, 66]. Complexity may be partly captured in models by relating measured environmental parameters to conditions directly experienced by organisms, such as an organism's body temperature (as opposed to the air temperature in most gridded datasets). Additionally, this complexity may be captured by modeling environmental effects based on developmental, physiological, or behavioral first principles, such as energy budget [67] or stem hydraulics models [68]. A mechanistic understanding of these responses might allow prediction of responses to extreme events that are rarely observed even in long-term observational studies.

A major challenge is that these approaches require many detailed observations or background information on the most important aspects of the environment for organisms. For many species and populations this information currently does not exist and will require extensive study of phenology, physiology, or behavior. Even in crops, which are among the most well-studied systems from an ecophysiological and developmental perspective, models struggle to effectively predict across multiple dimensions of novel environmental conditions [49]. Additionally, the fine scale data required to describe specific aspects of environment that impact fitness might be challenging to collect, model, and predict at large scales [61].

2.4 Biotic interactions

Studies of population response to the abiotic environment and forecasts of large scale dynamics typically overlook biotic interactions even though they can have large effects on individual fitness and population growth. For example, species with strong interactions (e.g., host-parasite, predator-prey) might have tightly coupled population dynamics, obscuring the effects of fluctuating abiotic conditions. Additionally, a given species might show strong population responses to local gradients in community composition, as in the case of an early or late successional specialist. Biotic constraints such as resource competition may limit organisms' ability to take advantage of favorable conditions, such as greater resource supply (e.g., higher rainfall for plants) [69, 70]. Abiotic conditions may determine the na-

ture of species interactions, such as when stress shifts the balance between competition and facilitation [71]. The effects of abiotic change on populations may be constrained and mediated by these biotic interactions.

Incorporating biotic interactions in predictive models is challenging because of the high dimensionality of community ecology (i.e., the number of species and higher order interactions among them) and the potential for no-analog future assemblages (Figure 1). When species interactions have a strong influence on population growth, shifts in community composition caused by abiotic environmental change can offset or even overwhelm the direct effects of abiotic change [5, 72, 73]. Unraveling these direct and indirect effects would be straightforward if the effects of abiotic factors and biotic interactions on demographic rates were independent and additive, but we have good reasons to expect complex interactions. For example, abiotic change might reduce fitness of a superior competitor more than an inferior competing species, leading to a population increase of the inferior competitor despite less favorable abiotic conditions [74]. Studies often rely on interannual climate fluctuations to observe abiotic-biotic interactions, but when future abiotic conditions are novel, communities may also enter novel states, thwarting prediction of a focal species' population dynamics based on recent observations (Figure 1). Furthermore, direct human impacts on populations will generate novel communities and biotic interactions [75]. The potential for such interactions illustrates the need for experiments that manipulate both abiotic conditions and biotic interactions.

Some researchers have fit joint models of community members' responses to abiotic conditions while including interactions among species [4, 76, 77]. However, such an approach requires large amounts of data, and it remains unclear how these efforts will succeed at forecasting under environmental change when models are fit to noisy observational data [78, 79]. A major reason there has been slow empirical progress in this area is the challenge in experimentally manipulating multiple community members (a problem of high dimensionality). Mechanistic models of species interactions are one potential way to include these interactions in predictive models [5, 80]. Alternate approaches focus on forecasts of community- or ecosystem-wide responses to environmental change [2]. These aggregate over the many dimensions of community variation, but may offer little information about dynamics of individual species of interest.

2.5 Demography

Simple, linear, and phenomenological models that relate environment to abundance might work well for organisms with very simple life cycles. For organisms with more complex life cycles, ignoring differences in how individuals of different ages, life stages, sizes or sexes respond to the environment may be problematic (Figure 1). When these sources of heterogeneity are important, models need to account for how individual vital rates (e.g., survival, growth, and fecundity) vary as a function of both demographic state (e.g., age, size, sex) and the environment [81].

To appreciate why accounting for the interactive effects of population structure and the environment is important, consider a phenomenological approach that relates local abundance of a long-lived plant directly to an environmental driver. If the population is well-established, it is likely dominated by large, mature individuals. Year-to-year variation in abundance will reflect the growth and mortality of these large individuals, and contributions from rare recruitment events will be small. If the mature individuals are stress-tolerant, we might find only a weak correlation between drought and population growth rate. Extrapolating this relationship to project the potential impacts of increasing aridity under climate change would predict minimal impacts on the population. But what if seedlings are less stress tolerant, and recruitment is limited to infrequent cool, wet years [82]? After disturbance of the established stand, the population might have difficulty recovering under future, more arid, conditions. In contrast, a model that accounts for interactions between size or stage structure and the environment could correctly predict both the limited impacts of increasing aridity on a population dominated by mature individuals, as well as a reduced ability of the population to recover following disturbance (when population growth depends most on recruitment). This hypothetical example shows how traditional demographic analyses, such as matrix population models [83] and integral population models [84], can

handle both high dimensionality (interactions between the environment and multiple vital rates) and novel conditions (changes in stage structure and environment).

While the existing tools for modeling demography and population growth are sophisticated, they have almost always been applied to just one local population at a time. Scaling up to multiple populations along complex environmental gradients represents a much greater challenge. Fortunately, many demographers are now focused on this task, using approaches described as "landscape demography" [85–87] or "dynamic range models" [88–90]. These approaches incorporate differences among populations in environmental context and stage-structure, but have largely ignored genetic or phenotypic variation among populations and species interactions.

3 How do we determine which of these processes are most important for prediction?

3.1 Including all processes in a model may not improve prediction

We have reviewed compelling arguments for why each of the processes described in Section 2 must be considered in a regional-scale population projection. However, a critical point that this body of literature has ignored is that including all of these processes in a model may not improve prediction. We frame the problem around regional population forecasting for a given species. However, regional models may be built for additional response variables of interest, such as population genetic composition, community diversity, and ecosystem function [4, 49] that may require added complexity.

Uncertainty in predictive modeling can be divided into two components: bias, caused by incorrect and overly simple models, versus variance, due to low numbers of observations relative to the number of estimated parameters [91]. While it may appeal to biologists to create models that incorporate ever more processes in greater realism, increasing model complexity can increase errors resulting from poorly estimated parameters, lead to overfitting, and decrease interpretability (making it harder to troubleshoot models). Optimal models for population prediction must balance bias against variance, and optimal models are often surprisingly simple [92, 93]. These problems have already presented themselves in population forecasting, but there has not been a reckoning for the complex models that would be the logical extension of the current movement to incorporate all the processes in Section 2.

3.2 Comparing models

Optimal predictive models are typically selected through empirical comparison or regularization (i.e., reducing complexity of a more general model; [94]). While we may have extensive knowledge about some components of the system, we may lack empirical results to constrain certain model formulations. Thus, we must make predictions with a set of models that represent different processes, test those predictions using proper and local scores based on independent data (data not used for model fitting; [95]), assess which model works best, and rigorously analyze uncertainties and biases [96].

To date, many efforts to predict species distributional changes or regional population dynamics have been conducted without out-of-sample validation. Without out-of-sample validation, we cannot determine how well predictions will fare under the novel conditions of the future. Experiments that manipulate environments to represent potential future conditions will provide important information. By testing model predictions locally (i.e., for specific locations in geographic or parameter space), we can begin to understand where and why models fail. This kind of validation and model comparison has yet to be performed for a regional-scale population model that incorporates all of the categories of processes described above, meaning that the relative importance of these processes is unknown, as is the utility of an integrative model.

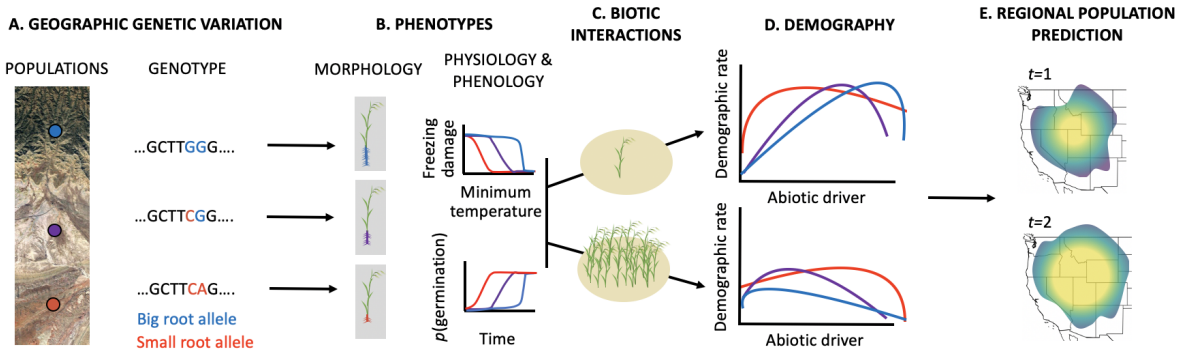

Figure 2: Reaction norms (figures with colored curves) occupy a middle position in a hierarchy of biological complexity among the mechanisms discussed here. Thus reaction norms can be viewed as a conceptual basis for building integrative models of regional population dynamics. We show a hypothetical plant example. Predicting regional dynamics (E) may require understanding (A) geographic patterns of genetic variation in environmental response, (B) the phenotypes involved and complex environment-performance relationships, such as physiological thresholds, (C) biotic interactions such as intraspecific density dependence, or (D) individual demographic rate responses to environment.

3.3 Reaction norms as a conceptual basis for integrative models

Traditionally, each of the important candidate processes described above has been studied separately. To test their relative value for predictions, we need new approaches for integrating information about different processes in statistical models. Without an integrative model, we cannot compare models incorporating more than one process, let alone consider interactions among these processes. For example, information about a genetic variant or a physiological threshold that affects organisms of a particular size class might be far more useful in a size-structured demographic model than in a model that simply tracks population density.

Reaction norms, which describe the change in a trait or performance across environments, provide a framework to integrate the processes discussed in the preceding sections. Reaction norms occupy a middle position in a hierarchy of biological complexity among the mechanisms we discussed, and thus they may be viewed as a nexus for integration (Figure 2). It is the complexity of these reaction norms that unites the arguments for incorporating mechanism into population forecasts (Section 2, Figure 1). We suggest that reaction norms are an intuitive concept that can be embedded in integrative population models. Genetic variation can be linked to reaction norms of lower-level traits like phenology, behavior, and physiology, which can then be linked to reaction norms for individual demographic rates, and fitness (Figure 2). However, previous research has described the response of one vital rate or trait to just one abiotic driver measured under controlled conditions. Such idealized reaction norms may be of little use for predicting population dynamics at a single location affected by many interacting environmental drivers, let alone for scaling up population models across a heterogeneous region containing genetically differentiated populations. The complexities we highlight require higher dimensional reaction norms (Figure 1).

One benefit of shifting to models that include reaction norms for key traits is that population models are more easily linked to ecosystem processes. For example, if photosynthetic rates are built into population models of a plant, it is easier to build integrated population-ecosystem models that allow us to assess how environmental changes jointly impact both (e.g., as has been investigated in crop models [49]). The complexity of trait relationships with demography and ecosystem processes requires a substantial integration of underlying mechanisms into population-ecosystem models [97].

4 Integrative hierarchical framework

We propose a hierarchical Bayesian framework to construct an integrated reaction norm model with genetic, phenotypic, and demographic components (Figure 3). These components correspond to multiple types of experimental and observational data, including DNA sequences, phenotypes, vital rates, and environmental conditions. The components of this model are linked by relationships that have been historically modeled separately. To account for and link the processes described above, an integrative model is required ([98] chapter 25). We propose straightforward extensions and generalizations of past efforts that enhance the power of our inferences and compare the importance of different components, and build on approaches developed for other integrated models that borrow strength from multiple data sources [99, 100].

We briefly describe an integrated hierarchical model that is comprised of three levels (Figure 3): genetic, phenotypic, and demographic. We offer more detail in the following sections. Our example model connects the three levels using conditional stochastic models with latent processes that depend on each other. In this case, the natural scaling between levels implies the conditional structure with the demographic process depending on the phenotypic process, which, in turn, depends on the genetic process.

The genetic component allows allele frequencies to change in space and time as a function of underlying environmental selective gradients and gene flow. This part of our approach is similar to genotype-environment association models that identify loci locally adapted to specific environments [63, 101], or spatial models of allele frequency turnover [102].

The genetic component is linked to a phenotypic model that determines how genotype and environment determine phenotypes. This part of the model captures processes usually studied by genetic mapping approaches that identify genetic loci causing trait variation [103], or genomic selection approaches that model the whole-genome contribution to quantitative trait variation [104], including models of trait variation across environments [45, 48]. This part of the model also captures plastic trait responses to environment (trait reaction norms) that are often a topic in ecophysiology, behavioral ecology, and functional ecology.

Finally, the demographic component determines how traits influence vital rates and ultimately fitness. This approach mirrors evolutionary ecology, quantitative genetic analyses of how trait variation is related to vital rates [105], changes in these relationships across environments (i.e., how traits mediate reaction norms of demography) [106, 107], as well as how selection may change across life stages/vital rates [108]. These model components cover the processes discussed in Section 2 and their integration is described below.

4.1 Modeling Genetic Processes based on Environment and Space

We start with an approach to model change in allele frequency along environmental gradients and in space in a way that facilitates prediction of allelic state at genetic markers (e.g., single nucleotide polymorphisms, SNPs) for a given location. Obtaining the predictive distribution of allele frequencies allows us to make inference on changes in allele frequencies under changing environmental conditions or with gene flow [110]. Parametric models allow us to properly account for, and learn about, uncertainty in a way that can be correctly propagated through to other types of inference. To generalize the concept of genotype-environment associations in a way that accounts for individual-level variability and mechanistic sources of gene flow, we use a hierarchical modeling structure where the individual-based genotype data g_{jil} for individual $i = 1, \ldots, n_j$ in population $j = 1, \ldots, J$ and for genetic locus $l = 1, \ldots, L$, arise stochastically as

$$g_{jil} \sim \text{Bern}(z_{jl})\,. \tag{1}$$

The population-level allele frequency is modeled as

$$f(z_{jl}) = \mathbf{x}_j'\boldsymbol{\beta}_l + \eta_{jl}\,, \tag{2}$$

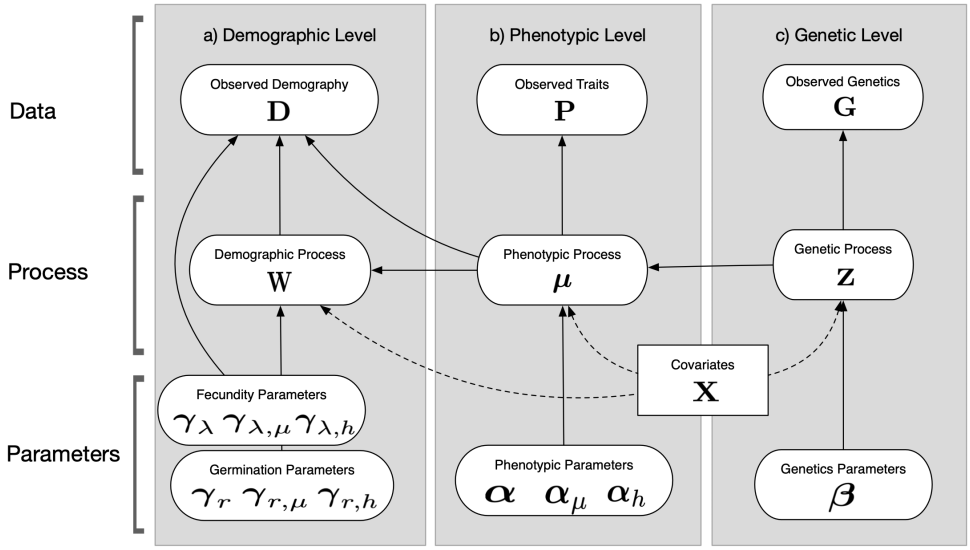

Figure 3: The directed acyclic graph [e.g., 109] for our integrated reaction norm model. Solid and dashed arrows indicate stochastic and deterministic connections, respectively. Parameter descriptions in full can also be found in supplemental Tables 1 and 2.

where $\mathbf{x}_j$ represent a set of covariates (e.g., topography, soil type, precipitation) and f is an appropriate link function. In this model (2), the coefficients $\boldsymbol{\beta}_l$ connect the environmental variables to allele frequency at SNP l in a way that accounts for local adaptation.

In cases where we need to account for patterns of gene flow and population structure, the η_{jl} are random effects (e.g., spatially-structured) [110–112] that we can model jointly as

$$\boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}) \,, \tag{3}$$

where the covariance matrix is parameterized as

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{R} \,. \tag{4}$$

The $J \times J$ matrix $\mathbf{R}$ may be specified as an intrinsic conditional autoregressive (ICAR; [113]) correlation matrix that depends on geographic information about the environment. Hanks and Hooten (2013) showed that this type of ICAR formulation provides a formal statistical model for gene flow based on circuit theory. By allowing the random effects η_{jl} in (2) to be spatially structured, we can accommodate gene flow mechanisms such as isolation by barrier, distance, and resistance [112]. Note that historical patterns of gene flow, especially during colonization when drift is strong, may differ from current gene flow, and the model can be generalized to accommodate such processes.

In the presence of temporal data, the model could be formulated as

$$g_{jilt} \sim \mathrm{Bern}(z_{jlt}) \,, \tag{5}$$

for time $t = 1, \ldots, T$, where the time-specific population-level allele frequency is modeled as

$$f(z_{jlt}) = \mathbf{x}'_{jt} \boldsymbol{\beta}_{lt} + \eta_{jlt} \,. \tag{6}$$

This extended model formulation explicitly accounts for changes in the environment over time to help us understand how environment aids in the prediction of genotypes. The model is generative and can be used to simulate genetic data to better understand the effects of future environments on genotypes.

This modeling framework can also be generalized to account for multilocus dependence explicitly. To do that, we introduce another random effect to the model in (2):

$$f(z_{jl}) = \mathbf{x}_j'\boldsymbol{\beta}_l + \eta_{jl} + \xi_{jl}\,, \tag{7}$$

where, jointly across loci, the random effects are assumed to be structured according to their linkage distance (if on the same chromosome), such that $\boldsymbol{\xi}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$. Not accounting for multilocus associations when they exist can obscure our inference about the importance of physically linked loci. Furthermore, multilocus dependence could improve prediction of genotypes for cases where the genotype is only partially observable (e.g., because of limited resources and genetic markers).

4.2 Incorporating Phenotypes

We can integrate genotype-phenotype and environment-phenotype relationships into our model. Consider the measurement of p_{jiq} for trait q on individual i in population j. To connect the phenotype and genotype in a statistical model, let

$$p_{jiq} \sim [p_{jiq}|\mu_{jiq}, \boldsymbol{\theta}_q]_q\,, \tag{8}$$

such that $[p_{jiq}|\mu_{jiq}, \boldsymbol{\theta}_q]_q$ is a trait-specific conditional data model with $\boldsymbol{\theta}_q$ as trait-specific parameters. We let the trait data model vary to accommodate a variety of traits; for example, for binary traits, the distribution in (8) should be Bernoulli and $\boldsymbol{\theta}_q = \emptyset$.

We now introduce the first reaction norm presented in our model, describing plastic trait change across environments (e.g., Figure 2 "Phenotypes"), as well as a way to accommodate higher dimensionality due to genetic variation in reaction norms. The mean μ_{jiq} of the conditional trait distribution is modeled as

$$f(\mu_{jiq}) = \mathbf{x}_j'\boldsymbol{\alpha} + \mathbf{z}_{ji}'\boldsymbol{\alpha}_\mu + h(\mathbf{x}_j, \mathbf{z}_{ji}, \boldsymbol{\alpha}_h) + \varepsilon_{jiq}\,, \tag{9}$$

where f is a suitable link function and $\mathbf{x}_j'\boldsymbol{\alpha}$ accounts for phenotypic plasticity (i.e., environmental effects on traits), $\mathbf{z}_{ji}'\boldsymbol{\alpha}_\mu$ accounts for genetic effects on the trait, $h(\mathbf{x}_j, \mathbf{z}_{ji}, \boldsymbol{\alpha}_h)$ is a function (e.g., a set of interactions) that accounts for genetic variation in plasticity (i.e. reaction norms), and the error term is correlated by a relevant grouping structure due to genetic similarity among individuals or latent spatial structure, for example. The genetic component of traits changes in space and time due to changes in allele frequency as described in the previous section. Jointly across individuals and groups, the errors are modeled as

$$\boldsymbol{\varepsilon}_q \sim \mathrm{N}(\mathbf{0}, \sigma_q^2\mathbf{I} + \boldsymbol{\Sigma}_y)\,, \tag{10}$$

where the variance component σ_q^2 may be set to zero as necessary for certain traits q.

Up to this point, we have described what might be termed an "integrated landscape genomic-reaction norm model," following common usage of terms. We can generalize this model to account for temporally indexed trait data like the genotype model described in the previous section. Finally, the modular nature of the two models implies that we may be able to apply recursive Bayesian techniques to fit the model in a computationally efficient framework [114, 115]. When genetic data are not available or desired, patterns of local adaptation in genetic trait-environment correlations can be accommodated in our framework by treating the genetic process in section 4.1 as latent.

4.3 Incorporating Demography

We can learn about population dynamics using demographic models, the specific form of which will vary depending on the study system. For example, at site j we observe n_j individual-level survival outcomes ($w_{ij} = 1$) that we model as $w_{ij} \sim \mathrm{Bern}(r_{ij})$, where r_{ij} represents the survival probability of individual i. By specifying individual-level demography, we can link individual genotype and phenotype to environment and the effects on demography. Our approach is to build reaction norms of demographic vital rates that depend on (possibly latent) phenotype-environment interactions, where the phenotypes themselves are modeled as reaction norms as described in the previous section.

At the end of a reproductive season, we can model the number of offspring of a hermaphroditic individual i, d_{ij}, as arising from the zero-inflated distribution

$$d_{ij} \sim \begin{cases} \text{Pois}(\lambda_{ij}) & , w_{ij} = 1 \\ 0 & , w_{ij} = 0 \end{cases}, \tag{11}$$

where λ_{ij} represents the fecundity of individuals surviving to reproduction.

We seek to infer the effect of phenotype and environment on individual-level fecundity and survival at site j. Consider the latent process model for individual-level survival probability r_{ij} where

$$f(r_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\gamma}_r + \boldsymbol{\mu}'_{ij}\boldsymbol{\gamma}_{r,\mu} + h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{r,h}), \tag{12}$$

and the covariates and associated regression coefficients are specified in the same manner as in the genetic and phenotypic model components. This model allows us to account for multivariate phenotypic effects $\boldsymbol{\mu}_{ij}$ on survival. When important traits are unknown or unmeasured (a common occurrence) we may use genotypes in place of, or in addition to $\boldsymbol{\mu}_{ij}$, to model selection on latent traits. We can specify a similar model for fecundity λ_{ij} as

$$\log(\lambda_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\gamma}_\lambda + \boldsymbol{\mu}'_{ij}\boldsymbol{\gamma}_{\lambda,\mu} + h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{\lambda,h}). \tag{13}$$

Biotic interactions (e.g., conspecific density dependence, parasitism) can be incorporated via $\mathbf{x}_{ij}$. This model can be fit from a hierarchical perspective (perhaps jointly with the previously described genetic and phenotypic components).

To review, we have a genetic model component connected to a phenotypic component, which is then linked to selection and demography. Note that the $h(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{\gamma}_{\lambda,h})$ function in (13) (and the similar function in (12)) and the $\mathbf{x}'_j\boldsymbol{\beta}_l$ term in (2) both represent changes in selection across environmental gradients $\mathbf{x}_{ij}$, and the information on these changes in selection arises from both the distribution of alleles across environments, their effects on phenotypes, and the changes in selection on these phenotypes across environments [116]. Together with the genetic and phenotypic model components, we can express the full integrated hierarchical genetic-phenotypic-demographic-environmental model as a directed acyclic graph (Figure 3).

We can generalize the ecological model component by allowing for additional data sources when available. For example, we can fuse presence-only and occupancy (with no individual identification) data, as well as unmarked counts of individuals if replicates are available, by adapting (2) to include the point process, occupancy, or N-mixture model frameworks [98]. Furthermore, we can accommodate time series data by indexing the data with a t subscript as described in Section 4.1.

4.4 Resulting Inference

The full hierarchical model in Figure 3 is generative and can be used to simulate data to better understand the effects of future environmental regimes on populations, phenotypes, and genotypes. For example, by comparing contemporary genotypes in a given location with genotypes under future environments, we can better understand sensitivity to changes in environment [15] and link this inference to spatio-temporal predictor variables and create maps of these sensitivities with uncertainty.

Each model component in Figure 3 can be fit individually to the relevant subset of data while conditioning on known or hypothesized elements from other components. However, the benefit of constructing a jointly specified integrated model is that we can fit the full hierarchical model and simultaneously learn about all the model quantities while allowing for feedbacks to occur among demographic, phenotypic, and genetic processes. The full hierarchical model also allows the uncertainties to propagate among model components so that we can make valid conclusions about the relative importance of each process. We can also compare predictions from the full model with predictions from simpler models that ignore one or more components and the processes they represent. These comparisons will identify the optimal level of complexity for prediction. When implementing the model, if additional flexibility

is needed, we can utilize basis function expansions of the environmental, genetic, and phenotypic space [117] to account for nonlinearities in the relationships.

To optimize predictive ability, we can hold out data from a subset of sites (or time points) for model validation. We can use the integrated model to obtain the posterior predictions for the latent processes. The direct comparison of observed and predicted population growth is essential for quantifying sources of uncertainty in the model, an important step in guiding future research to improve the model.

4.5 Feasibility

Fitting our integrated model will require an unprecedented, coordinated data collection effort. But unprecedented does not mean impossible. Annual plants may be a good starting point because their simple life cycles facilitate estimation of lifetime fitness, even in experimental settings. For example, existing knowledge and resources make the model annual plant *Arabidopsis thaliana* amenable [40, 47]. We are leading a network of researchers tackling the goal of this paper using *Bromus tectorum*, a widely distributed annual grass invasive in western North America. Even some long-lived perennial plants have been studied using common gardens, demographic observations, and genomic techniques (especially in species of economic value) [118, 119]. While controlled field or laboratory experiments may even be possible for some small animals [120], for many species controlled experiments are extremely difficult. Additionally, the complexity of biological systems may resist our efforts to experiment, observe, and build mechanistic models. In the face of such challenges, integrative regional forecasting models may be built with fewer of the processes described above, or with phenomenological components. The strength of our suggested approach is its ability to quantify the predictive value of any processes that can be feasibly modeled.

5 Conclusion

It is clear that a wide range of biological processes can be incorporated into large scale forecasts of population dynamics under environmental change. But the route to accurate prediction might not involve building models of ever-increasing complexity. While the integrated model we describe does account for many underlying processes, we emphasize that this full model is intended for development and may not itself be the ideal model for prediction. To advance the field of regional population prediction we need extensive empirical study and an iterative process of model improvement [121]. Which processes are important for good forecasting may depend on the time-scale of the forecast (or forecast horizon) [122].

We presented a statistical framework to integrate these diverse processes in a model for use in prediction. This framework is novel because it formally unifies genetic, phenotypic, and demographic components into an integrated reaction norm model and suggests general computational methods for fitting it. We can compare our methods with other existing, less integrated approaches and quantify the value added when formally borrowing strength across multiple sources of data in a single, cohesive hierarchical modeling framework. The answer to the question we posed "what processes must we understand to forecast regional scale population dynamics?" will be found only through careful study that generates the extensive multi-dimensional data discussed here combined with an integrative modeling approach.

Acknowledgements

The authors thank Kristen Ruegg, Erika Zavaleta, and the BromeCast Network for early discussions and insights related to this work, and Samuel Scarpino for comments on the manuscript. Funding was provided by NSF DEB 1927177, 1927282, and 1927009. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

(1) Boutillier, P. et al. *Bioinformatics* **2018**, *34*, Publisher: Oxford Academic, i583–i592.

(2) Medvigy, D. et al. *Journal of Geophysical Research: Biogeosciences* **2009**, *114*, G01002.

(3) Bay, R. A. et al. *The American Naturalist* **2017**, *189*, 463–473.

(4) Clark, J. S. et al. *Ecological Applications* **2014**, *24*, 990–999.

(5) Buckley, L. B. *Annals of the New York Academy of Sciences* **2013**, n/a–n/a.

(6) Kearney, M.; Porter, W. *Ecology Letters* **2009**, *12*, Publisher: John Wiley & Sons, Ltd, 334–350.

(7) Urban, M. C. et al. *Science* **2016**, *353*, Publisher: American Association for the Advancement of Science Section: Review, DOI: 10.1126/science.aad8466.

(8) Dietze, M. C. et al. *Proceedings of the National Academy of Sciences* **2018**, 201710231.

(9) Dawson, T. P. et al. *Science* **2011**, *332*, Publisher: American Association for the Advancement of Science Section: Review, 53–58.

(10) Clark, J. et al. *Science* **2001**, *293*, 657–660.

(11) Guisan, A.; Thuiller, W. *Ecology Letters* **2005**, *8*, Publisher: John Wiley & Sons, Ltd, 993–1009.

(12) Crone, E. E. et al. *Conservation Biology* **2013**, *27*, Publisher: John Wiley & Sons, Ltd, 968–978.

(13) Harris, D. J. et al. *PeerJ* **2018**, *6*, DOI: 10.7717/peerj.4278.

(14) Merow, C. et al. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, E3276–E3284.

(15) Capblancq, T. et al. *Annual Review of Ecology, Evolution, and Systematics* **2020**, *51*, _eprint: https://doi.org/10.1146/annurev-ecolsys-020720-042553, null.

(16) Crozier, L.; Dwyer, G. *The American Naturalist* **2006**, *167*, Publisher: The University of Chicago Press, 853–866.

(17) Helmuth, B. et al. In *Annual Review of Physiology*, WOS:000228406800009; Annual Reviews: Palo Alto, 2005; Vol. 67, pp 177–201.

(18) Hooten, M. B. et al. *Ecology Letters* **2019**, *22*, 377–389.

(19) Buckley, L. B. et al. *Ecology Letters* **2010**, *13*, 1041–1054.

(20) Hefley, T. J. et al. *Ecology Letters* **2017**, *20*, 640–650.

(21) Moran, E. V.; Alexander, J. M. *Ecology Letters* **2014**, *17*, Publisher: John Wiley & Sons, Ltd, 637–649.

(22) Williams, J. W.; Jackson, S. T. *Frontiers in Ecology and the Environment* **2007**, *5*, Publisher: John Wiley & Sons, Ltd, 475–482.

(23) Scarpino, S. V.; Petri, G. *Nature Communications* **2019**, *10*, Number: 1 Publisher: Nature Publishing Group, 898.

(24) Pennekamp, F. et al. *Ecological Monographs* **2019**, *89*, Publisher: John Wiley & Sons, Ltd, e01359.

(25) Hébert-Dufresne, L. et al. *Nature Physics* **2020**, *16*, Number: 4 Publisher: Nature Publishing Group, 426–431.

(26) Reichert, P.; Omlin, M. *Ecological Modelling* **1997**, *95*, ISBN: 0304-3800 Publisher: Elsevier, 289–299.

(27) Clark, J. S. et al. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2012**, *367*, 236–246.

(28) Wagner, C. H. *The American Statistician* **1982**, *36*, Publisher: Taylor & Francis _eprint: https://www.tandfonline.co 46–48.

(29) Piantadosi, S. et al. *American Journal of Epidemiology* **1988**, *127*, Publisher: Oxford Academic, 893–904.

(30) Peterson, M. L. et al. *Global Change Biology* **2019**, *In press*, DOI: 10.1111/gcb.14562.

(31) Josephs, E. B. *New Phytologist* **2018**, *219*, Publisher: John Wiley & Sons, Ltd, 31–36.

(32) Hoffmann, A. A.; Harshman, L. G. *Heredity* **1999**, *83*, Number: 6 Publisher: Nature Publishing Group, 637–643.

(33) Fournier-Level, A. et al. *Science* **2011**, *334*, 86–89.

(34) Carius, H. J. et al. *Evolution* **2001**, *55*, Publisher: John Wiley & Sons, Ltd, 1136–1145.

(35) Cousyn, C. et al. *Proceedings of the National Academy of Sciences* **2001**, *98*, Publisher: National Academy of Sciences Section: Biological Sciences, 6256–6260.

(36) Torres-Dowdall, J. et al. *Evolution* **2012**, *66*, Publisher: John Wiley & Sons, Ltd, 3432–3443.

(37) Schiffers, K. et al. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2012**, *368*, DOI: 10.1098/rstb.2012.0083.

(38) Bay, R. A. et al. *Science* **2018**, *359*, 83–86.

(39) Razgour, O. et al. *Proceedings of the National Academy of Sciences* **2019**, *116*, ISBN: 9781820663118 Publisher: National Academy of Sciences Section: Biological Sciences, 10418–10423.

(40) Exposito-Alonso, M. et al. *Nature Ecology & Evolution* **2018**, *2*, 352.

(41) Rhoné, B. et al. *Nature Communications* **2020**, *11*, Number: 1 Publisher: Nature Publishing Group, 5274.

(42) Angert, A. L. et al. *Integrative and Comparative Biology* **2011**, *51*, Publisher: Oxford Academic, 733–750.

(43) Ikeda, D. H. et al. *Global Change Biology* **2017**, *23*, Publisher: John Wiley & Sons, Ltd, 164–176.

(44) Wang, T. et al. *Ecological Applications* **2010**, *20*, Publisher: John Wiley & Sons, Ltd, 153–163.

(45) Tiezzi, F. et al. *Journal of Dairy Science* **2017**, *100*, 2042–2056.

(46) Lasky, J. R. et al. *Science Advances* **2015**, *1*, e1400218.

(47) Gienapp, P. et al. *Trends in Ecology & Evolution* **2017**, *32*, 897–908.

(48) Messina, C. D. et al. *European Journal of Agronomy* **2018**, *100*, 151–162.

(49) Peng, B. et al. *Nature Plants* **2020**, *6*, Number: 4 Publisher: Nature Publishing Group, 338–348.

(50) Rodgers-Melnick, E. et al. *Proceedings of the National Academy of Sciences* **2016**, *113*, E3177–E3184.

(51) Lasky, J. R. et al. *Molecular Biology and Evolution* **2014**, *31*, 2283–2296.

(52) Jones, F. C. et al. *Nature* **2012**, *484*, 55–61.

(53) Polechová, J. et al. *The American Naturalist* **2009**, *174*, E186–E204.

(54) Aitken, S. N. et al. *Evolutionary Applications* **2008**, *1*, 95–111.

(55) Peischl, S. et al. *Molecular Ecology* **2013**, *22*, 5972–5982.

(56) Gienapp, P. et al. *Evolution Letters* **2019**, *3*, 142–151.

(57) Rudman, S. M. et al. *Nature Ecology & Evolution* **2018**, *2*, Number: 1 Publisher: Nature Publishing Group, 9–15.

(58) Whittaker, R. H. et al. *The American Naturalist* **1973**, *107*, ArticleType: research-article / Full publication date: May - Jun., 1973 / Copyright © 1973 The University of Chicago Press, 321–338.

(59) John, R. et al. *Proceedings of the National Academy of Sciences* **2007**, *104*, 864–869.

(60) Donohue, K. et al. *Trends in Ecology & Evolution* **2015**, *30*, 66–77.

(61) Buckley, L. B. et al. *Integrative and Comparative Biology* **2018**, *58*, Publisher: Oxford Academic, 38–51.

(62) Phillips, S. J. et al. *Ecological Modelling* **2006**, *190*, 231–259.

(63) Fitzpatrick, M. C.; Keller, S. R. *Ecology Letters* **2015**, *18*, 1–16.

(64) Hooten, M. et al. *International Statistical Review* **2020**, *88*, 441–461.

(65) Kearney, M.; Porter, W. *Ecology Letters* **2009**, *12*, 334–350.

(66) Lazarus, B. E. et al. *Journal of Ecology* **2018**, *106*, 571–585.

(67) Buckley, L. *The American Naturalist* **2008**, *171*, ArticleType: research-article / Full publication date: January 2008 / Copyright © 2008 The University of Chicago, E1–E19.

(68) Sterck, F. et al. *Proceedings of the National Academy of Sciences* **2011**.

(69) Lorts, C. M.; Lasky, J. R. *New Phytologist* **2020**, *n/a*, Publisher: John Wiley & Sons, Ltd, DOI: 10.1111/nph.16593.

(70) Farrior, C. E. et al. *The American Naturalist* **2013**, *181*, 314–330.

(71) Lortie, C. J.; Callaway, R. M. *Journal of Ecology* **2006**, *94*, 7–16.

(72) Gilman, S. E. et al. *Trends in Ecology & Evolution* **2010**, *25*, 325–331.

(73) Suttle, K. et al. *Science* **2007**, *315*, 640–642.

(74) Lasky, J. R. *Evolutionary Applications* **2019**, *12*, 1434–1448.

(75) Helmus, M. R. et al. *Nature* **2014**, *513*, Number: 7519 Publisher: Nature Publishing Group, 543–546.

(76) Ovaskainen, O. et al. *Methods in Ecology and Evolution* **2016**, *7*, Publisher: John Wiley & Sons, Ltd, 549–555.

(77) Clark, J. S. et al. *Proceedings of the National Academy of Sciences* **2020**, *117*, Publisher: National Academy of Sciences Section: Biological Sciences, 17074–17083.

(78) Freckleton, R. P. et al. *Journal of Animal Ecology* **2006**, *75*, Publisher: John Wiley & Sons, Ltd, 837–851.

(79) Detto, M. et al. *Ecology Letters* **2019**, *22*, Publisher: John Wiley & Sons, Ltd, 1923–1939.

(80) Uriarte, M. et al. *Ecology Letters* **2010**, *13*, 1503–1514.

(81) Tredennick, A. T. et al. *Ecology Letters* **2018**, *21*, Publisher: John Wiley & Sons, Ltd, 1757–1770.

(82) Niinemets, Ü. *Forest Ecology and Management* **2010**, *260*, 1623–1639.

(83) Caswell, H., *Matrix population models. Construction, analysis and interpretation. 2nd Edition.* Sinauer Associates, Inc. Sunderland, MA: 2001.

(84) Ellner, S. P. et al., *Data-driven Modelling of Structured Populations*; Lecture Notes on Mathematical Modelling in the Life Sciences; Springer International Publishing: Cham, 2016.

(85) Ehrlen, J.; Morris, W. F. *Ecology Letters* **2015**, *18*, 303–314.

(86) Gurevitch, J. et al. *The Quarterly Review of Biology* **2016**, *91*, 459–485.

(87) Gerber, B. D. et al. *Functional Ecology* **2019**, *33*, 1695–1707.

(88) Pagel, J.; Schurr, F. M. *Global Ecology and Biogeography* **2012**, *21*, 293–304.

(89) Schurr, F. M. et al. *Journal of Biogeography* **2012**, *39*, 2146–2162.

(90) Merow, C. et al. *Ecography* **2014**, *37*, 1167–1183.

(91) Geman, S. et al. *Neural computation* **1992**, *4*, ISBN: 0899-7667 Publisher: MIT Press, 1–58.

(92) Chevalier, M.; Knape, J. *Oikos* **2020**, *129*, Publisher: John Wiley & Sons, Ltd, 249–260.

(93) Ward, E. J. et al. *Oikos* **2014**, *123*, Publisher: John Wiley & Sons, Ltd, 652–661.

(94) Hooten, M. B.; Cooch, E. C. In *Quantitative Analysis in Wildlife Science*, Brennan, L. A. et al., Eds.; Johns Hopkins University Press: 2019.

(95) Hooten, M. B.; Hobbs, N. T. *Ecological Monographs* **2015**, *85*, 3–28.

(96) Conn, P. B. et al. *Ecological Monographs* **2018**, *88*, 526–542.

(97) Fridley, J. D. *Journal of Ecology* **2017**, *105*, Publisher: John Wiley & Sons, Ltd, 95–110.

(98) Hooten, M. B.; Hefley, T. J., *Bringing Bayesian Models to Life*; CRC Press: 2019.

(99) Schaub, M.; Abadi, F. *Journal of Ornithology* **2011**, *152*, 227–237.

(100) Hanks, E. M. et al. *Ecological Applications* **2011**, *21*, 1173–1188.

(101) Coop, G. et al. *Genetics* **2010**, *185*, 1411–1423.

(102) Yang, W.-Y. et al. *Nat Genet* **2012**, *44*, 725–731.

(103) Hirschhorn, J. N.; Daly, M. J. *Nature Reviews Genetics* **2005**, *6*, 95–108.

(104) VanRaden, P. et al. *Journal of Dairy Science* **2009**, *92*, 16–24.

(105) Lande, R.; Arnold, S. J. *Evolution* **1983**, *37*, 1210–1226.

(106) Wade, M. J.; Kalisz, S. *Evolution* **1990**, *44*, ArticleType: research-article / Full publication date: Dec., 1990 / Copyright © 1990 Society for the Study of Evolution, 1947–1955.

(107) Hall, M. C.; Willis, J. H. *Evolution* **2006**, *60*, Publisher: John Wiley & Sons, Ltd, 2466–2477.

(108) Schluter, D. et al. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **1991**, *246*, Publisher: Royal Society, 11–17.

(109) Hobbs, N. T.; Hooten, M. B., *Bayesian Models: A Statistical Primer for Ecologists*, Student; Princeton University Press: 2015.

(110) Hanks, E. M.; Hooten, M. B. *Journal of the American Statistical Association* **2013**, *108*, 22–33.

(111) Hanks, E. M. et al. *The Annals of Applied Statistics* **2016**, *10*, 1041–1062.

(112) Peterson, E. E. et al. *Ecological Monographs* **2019**, *89*, e01355.

(113) Ver Hoef, J. M. et al. *Ecological Monographs* **2018**, *88*, 36–59.

(114) Hooten, M. B. et al. *Environmetrics* **2016**, *27*, 322–333.

(115) Hooten, M. B. et al. *The American Statistician* **2020**, 1–10.

(116) Lasky, J. R. et al. *Molecular Ecology Resources* **2018**, *18*, 91–106.

(117) Hefley, T. J. et al. *Ecology* **2017**, *98*, 632–646.

(118) Yeaman, S. et al. *Science* **2016**, *353*, 1431–1433.

(119) Lowry, D. B. et al. *Proceedings of the National Academy of Sciences* **2019**, *116*, 12933–12941.

(120) Rudman, S. M. et al. *Proceedings of the National Academy of Sciences* **2019**, *116*, Publisher: National Academy of Sciences Section: PNAS Plus, 20025–20032.

(121) Ver Hoef, J. M.; Boveng, P. L. *The Journal of Wildlife Management* **2015**, *79*, 719–729.

(122) Adler, P. B. et al. *Ecography* **2020**, *n/a*, Publisher: John Wiley & Sons, Ltd, DOI: `10.1111/ecog.05271`.

Supplement

Variable	Description
d_{ij}	number of seeds per plant individual i at site j
λ_{ij}	fecundity of individual i at site j
w_{ij}	individual i germination outcome at site j
r_{ij}	individual i germination probability at size j
f	a suitable link function for regression (e.g., logit, probit, log)
$\mathbf{x}_j$	covariates (e.g., temperature, precipitation) at site j
$\boldsymbol{\mu}_{ij}$	individual i phenotype at site j
μ_{jiq}	phenotypic process associated with site j, individual i, and trait q
$\boldsymbol{\gamma}_r$	regression coefficients for the static, dynamic, and phenotypic aspects of the ecological component involving germination
$\boldsymbol{\gamma}_\lambda$	regression coefficients for the static, dynamic, and phenotypic aspects of the ecological component involving fecundity
$\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_\mu$, $\boldsymbol{\alpha}_h$	regression coefficients for the phenotypic model component
$\boldsymbol{\beta}_l$	regression coefficients for the static, dynamic, and genetic aspects of the genetic model component
$\boldsymbol{\Sigma}_y$	phenotypic covariance that can account for similarity among individuals and/or sites
σ^2_q	unstructured phenotypic variance for trait q
p_{jiq}	phenotypic data associated with site j, individual i, and trait q
$\mathbf{g}_{ji}$	aggregated genetic data associated with site j and individual i
g_{jil}	genetic data associated with site j for individual i at locus l (e.g., minor allele frequencies)
σ^2_g	spatial variance term associated with gene flow

Table 1: Description of variables in hierarchical model.

Index	Description
i	individual i
j	site j
q	trait q
l	locus or SNP l

Table 2: Description of indices for variables in hierarchical model.